

Correspondence Estimation from Non-Rigid Motion Information

Jonas Wulff^{a,b}, Thomas Lotz^b, Thomas Stehle^a, Til Aach^a and J. Geoffrey Chase^b

^aInstitute of Imaging and Computer Vision, RWTH Aachen University, Germany

^bCentre for Bioengineering, Department of Mechanical Engineering, University of Canterbury, Christchurch, New Zealand

ABSTRACT

The DIET (Digital Image Elasto Tomography) system is a novel approach to screen for breast cancer using only optical imaging information of the surface of a vibrating breast. 3D tracking of skin surface motion without the requirement of external markers is desirable. A novel approach to establish point correspondences using pure skin images is presented here. Instead of the intensity, motion is used as the primary feature, which can be extracted using optical flow algorithms. Taking sequences of multiple frames into account, this motion information alone is accurate and unambiguous enough to allow for a 3D reconstruction of the breast surface. Two approaches, direct and probabilistic, for this correspondence estimation are presented here, suitable for different levels of calibration information accuracy. Reconstructions show that the results obtained using these methods are comparable in accuracy to marker-based methods while considerably increasing resolution. The presented method has high potential in optical tissue deformation and motion sensing.

Keywords: Shape Reconstruction, Breast Cancer Screening, Digital Image Elasto-Tomography, Correspondence Estimation, Non-Rigid Deformation

1. INTRODUCTION

Breast cancer is the most common cancer type in women, accounting for about one fifth of cancer deaths.¹ Since mortality rates heavily depend on the cancer progress at the time of diagnosis, extensive screening programs have been established in most western countries, usually by means of X-ray mammography. The practice of regular, frequent screening, however, is not unquestioned. The main counter argument is that the radiation exposure might in some cases cause cancer itself. Besides the X-ray exposure, the heavy mechanical stress imposed on the breast is another shortcoming of mammography. Together, these factors lead generally to low compliance rates.²

The Digital Image Elasto Tomography (DIET) technology³ is a non-invasive elastographic screening approach that is based on the high stiffness contrast of a factor five to fifteen between cancerous and healthy breast tissue. Herein, the breast is actuated by a sinusoidally vibrating device, inducing oscillatory vibrations in the breast. The breast surface is then tracked through a single oscillation period, and the motion of the surface is analyzed. Additionally, the three-dimensional breast shape is reconstructed and used to simulate the surface motion for varying elastic property distributions within the breast. Comparing the real with the simulated surface motion gives an estimation of the presence, size, and location of a potential tumor. Alternatively, the surface motion can be used directly to estimate the presence of a tumor.⁴

To estimate the surface motion and reconstruct the breast shape, five calibrated cameras are positioned around the breast. The image processing then consists of two steps: *tracking* and *correspondence estimation*. In this context, *tracking* is the process of following a single point through the sequence (in this case, one period of steady-state motion). This step is necessary to determine the surface motion. *Correspondence estimation*, on the

Further author information: (Send correspondence to J. Geoffrey Chase)

Jonas Wulff: E-mail: jonas.wulff@rwth-aachen.de, Telephone: +49 (0) 241 8027 860

Thomas Lotz: E-mail: thomas.lotz@canterbury.ac.nz, Telephone: +64 (3) 364 2987 7384

Thomas Stehle: E-mail: thomas.stehle@lfb.rwth-aachen.de, Telephone: +49 (0) 241 8027 860

J. Geoffrey Chase: E-Mail: geoff.chase@canterbury.ac.nz, Telephone: +64 (3) 364 2987 7224

Til Aach: E-Mail: til.aach@lfb.rwth-aachen.de, Telephone: +49 (0) 241 8027 860

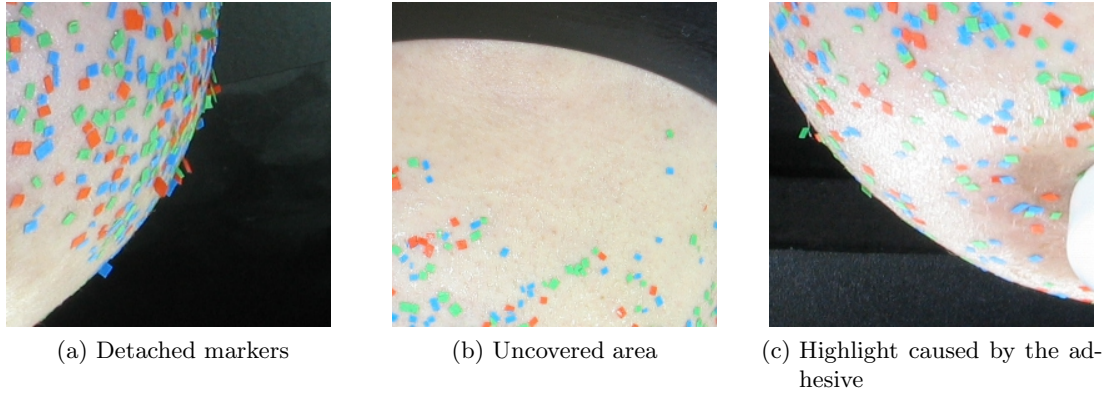


Figure 1: Examples for problems arising from the use of colored markers

other hand, refers to the detection of the same point in the views of multiple cameras at a given point in time. Correspondence estimation is required for the reconstruction of the breast shape, which in turn is necessary for the simulation.

In the original DIET approach, both processes are facilitated by small, colored markers. Using an adhesive, these markers are attached to the breast, covering the whole surface. With the markers, both tracking and correspondence estimation can be robustly achieved using nearest neighbor methods.⁵ The use of colored markers, however, introduces a number of problems:

- Due to the breast movement, markers can detach from the surface, potentially disturbing the image processing (Fig. 1(a)).
- Some areas of the breast, especially those near the inframammary fold, can remain unmarked, causing either holes or large areas of imprecision in the final triangulation (Fig. 1(b)).
- The adhesive causes highlights, overexposing certain breast areas, rendering image processing in these areas impossible (Fig. 1(c)).
- The procedure of attaching the adhesive and the markers and removing them afterwards is cumbersome for the patient.
- The achievable resolution is limited by the number of markers.

To overcome these problems, a method for both tracking and correspondence estimation using only skin texture was developed. Since the breast skin texture provides sufficient detail invariant to small motion, tracking is done by standard optical flow based methods.⁶

At a first glance, the geometric reconstruction resembles a *Structure-from-Motion* (SfM) problem. However, existing algorithms in this field (e.g.^{7,8}) usually assume established correspondences. These are used to determine camera calibration information - an approach that generalizes to problems such as determination of egomotion^{7,8} or calibration of intrinsic camera parameters.⁹ Even approaches that do not need explicit a-priori feature correspondences but establish correspondences in the process still rely on some sets of relatively sparse, distinct features such as corners.¹⁰

In the case presented here, the calibration information is known, whereas the feature correspondences are unknown. Direct feature matching using a feature descriptor such as SIFT¹¹ or SURF¹² is not possible, due to the heavy geometric distortions the features undergo in viewpoint changes between the cameras. Feature detectors that are invariant to some of these transformations exist, such as ASIFT,¹³ an affine-invariant feature detector. Preliminary tests conducted with some of these detectors, however, showed unacceptable performance.

Besides these approaches, the work done in the field of incorporating motion information to correspondence estimation is fairly sparse and, in general, restricted to rigid objects.¹⁴⁻¹⁶ For such constrained objects, motion

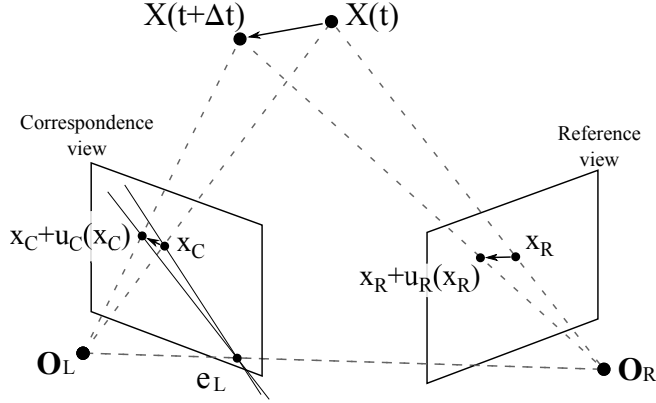


Figure 2: Epipolar setup.

information from several locations can be integrated into a single motion model for the whole object, an approach that is not feasible here. Thus, a correspondence estimation method was developed that is invariant to a change in an object’s visual appearance when observed from different viewing angles and explicitly utilizes motion information.

2. METHODOLOGY

Consider a usual epipolar setup as depicted in Fig. 2, consisting of two calibrated views, the *reference view* and the *correspondence view*. The relationship between points in both views, given in pixel coordinates, is defined by the fundamental matrix \mathbf{F} . Given a point $\mathbf{x}_R = (x_R, y_R, 1)^\top$ in homogeneous coordinates in the reference view, the corresponding point $\mathbf{x}_C = (x_C, y_C, 1)^\top$ in the correspondence view lies on the epipolar line defined by the epipolar constraint:

$$\mathbf{x}_C^\top \mathbf{F}^\top \mathbf{x}_R = 0 \quad (1)$$

As shown in Fig. 2, the epipolar constraint holds at any subsequent point in time, at which the points \mathbf{x}_R and \mathbf{x}_C have moved by the amounts $\mathbf{u}(\mathbf{x}) = (\Delta x(\mathbf{x}), \Delta y(\mathbf{x}), 0)^\top$ (indices R/C omitted for clarity). Defining $\mathbf{x}(t + \Delta t) = \mathbf{x}(t) + \mathbf{u}(\mathbf{x}(t))$, Eq. (1) has to hold not only for $\mathbf{x}_{R/C}$, but also for $\mathbf{x}_{R/C}(t + \Delta t)$ for all Δt :

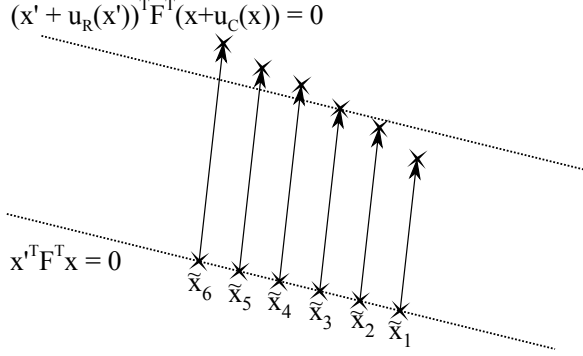
$$(\mathbf{x}_C(t + \Delta t))^\top \mathbf{F}^\top \mathbf{x}_R(t + \Delta t) = 0 \quad \forall t, \Delta t \quad (2)$$

Fig. 3(a) illustrates the actual matching process in a single frame, using a set of sampled points $\tilde{\mathbf{x}}_i$ along the epipolar line. In this case, $\tilde{\mathbf{x}}_3$ would be the best match. However, one problem with Eq. (2) is that due to small errors (most notably errors in the estimated motion and pixel sampling) it does not necessarily hold exactly true. Furthermore, the motion might match at multiple locations along the epipolar line. In Fig. 3(b) the error as given by Eq. (2) is shown against the possible position i (i.e. the sampling index as shown in Fig. 3(a)) along the epipolar line, emphasizing this ambiguity. These problems can be solved by integrating the matching information over time. Given a finite number T of frames (in the DIET context: one period of sinusoidal movement) and introducing a time and location dependent weighting factor $w(t, \mathbf{x}(t))$, the best correspondence $\tilde{\mathbf{x}}_C$ for a specified reference point \mathbf{x}_R can be found using

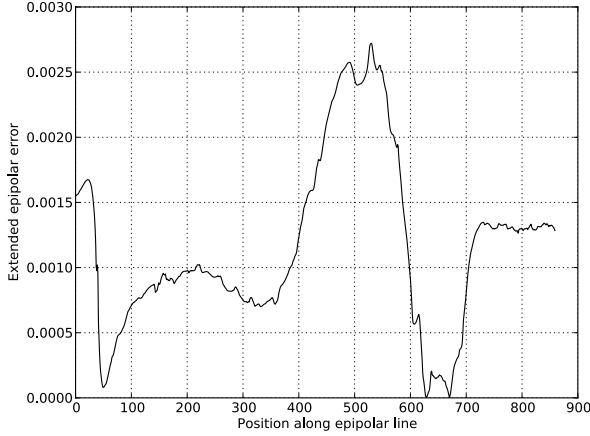
$$\tilde{\mathbf{x}}_C = \operatorname{argmin}_{\mathbf{x}_C} \sum_{t=1}^T w(t, \mathbf{x}_R(t)) |(\mathbf{x}_R(t))^\top \mathbf{F}^\top \mathbf{x}_C(t)| \quad (3)$$

Here, $\mathbf{x}(t)$ denotes an image point, and its tracked positions over time. The weighting factor $w(t, \mathbf{x}(t))$ is subject to the following constraints:

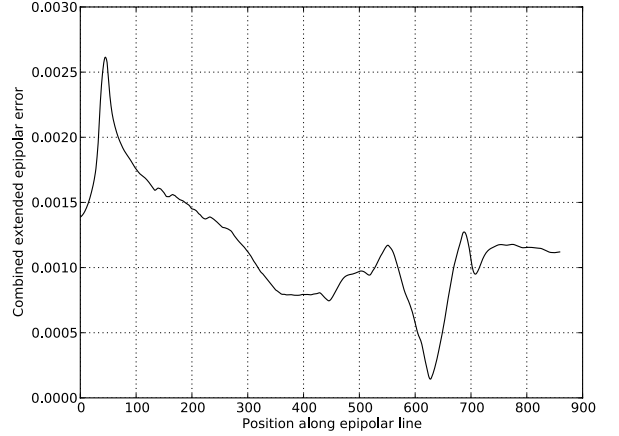
$$\begin{aligned} w(t, \mathbf{x}(t)) &\geq 0 \quad \forall t \in \{1, \dots, T\} \\ \sum_{t=1}^T w(t, \mathbf{x}(t)) &= 1 \end{aligned} \quad (4)$$



(a) Matching using the extended epipolar constraint



(b) Extended epipolar error, using a single frame



(c) Extended epipolar error, using a period

Figure 3: Matching process and extended epipolar errors using direct reconstruction

Thus $w(t, \mathbf{x}(t))$ weights the influence of the information for a given point $\mathbf{x}(t)$ inherent in the frame at time t . Considering Fig. 3(a) again, it can be seen that the more the motion around the match varies, the better this match can be estimated. This chicken-and-egg problem is solved by taking the fact into account that the point to be corresponded is given in the reference view. Therefore, the variation of motion surrounding that point in the reference view is a good estimate for the variation of motion surrounding the correct match in the correspondence view. Thus, given $\mathbf{u}_R(t, \mathbf{x}) = (\Delta x_R(t, \mathbf{x}), \Delta y_R(t, \mathbf{x}), 0)^\top$, $w(t, \mathbf{x}(t))$ can be estimated as

$$\tilde{w}(t, \mathbf{x}(t)) = \|\nabla [\Delta x_R(t, \mathbf{x}(t))]\|_2 + \|\nabla [\Delta y_R(t, \mathbf{x}(t))]\|_2 \quad (5)$$

Fulfilling condition (4), the final weights are written as

$$w(t, \mathbf{x}(t)) = \frac{\tilde{w}(t, \mathbf{x}(t))}{\sum_t \tilde{w}(t, \mathbf{x}(t))}. \quad (6)$$

Given sufficiently accurate calibration information, this approach already facilitates 3D reconstruction. It is henceforth referred to as *direct reconstruction*.

In order to generalize this approach to a wider range of applications beyond the DIET technology, it could be required to cope with distorted calibration information. A probabilistic approach to overcome these problems can be applied. In the theoretical case of a perfect match, instead of zero, Eq. (1) takes the finite value d_l (l for location), which is a function of \mathbf{x}_R and \mathbf{x}_C . Using the long form of Eq. (3) and further incorporating errors

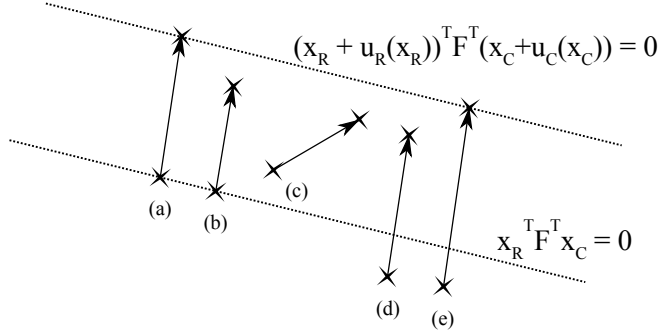


Figure 4: Possible distance cases

caused by an inaccurate motion estimation, an additional distance d_m (m for motion) can be defined as

$$d_m(\mathbf{x}_R, \mathbf{x}_C) = (\mathbf{x}_R + \mathbf{u}_R(\mathbf{x}_R))^T \mathbf{F}^T (\mathbf{x}_C + \mathbf{u}_C(\mathbf{x}_C)) - d_l(\mathbf{x}_R, \mathbf{x}_C) \quad (7)$$

The rationale of subtracting d_l in Eq. (7) is illustrated in Fig. 4: Case (a) shows a perfect match, case (b) a good match with respect to location (in this case, $d_l = 0$) but not with respect to motion, case (c) a bad match in both respects. Without including d_l in Eq. (7), (e) would result in a better match than (d). However, the motion of (d) is equal to the motion of the correct match (a) and should therefore be preferred. This can be achieved by virtually shifting case (d) by d_l , so that it falls on the epipolar line. Mathematically, without this shift, d_m would be influenced by both motion estimation and calibration errors, while d_l is only subject to the calibration error. Subtracting d_l thus effectively removes the calibration error from d_m . Assuming a Gaussian error on d_l , the probability p_l for a point to correspond to a given reference point with respect only to the corresponding point's location (hence the index l) is:

$$\tilde{p}_l(\mathbf{x}_C | \mathbf{x}_R) = \frac{1}{\sqrt{2\pi}\sigma_l} \exp\left(-\frac{d_l(\mathbf{x}_R, \mathbf{x}_C)^2}{2\sigma_l^2}\right) \quad (8)$$

$$p_l(\mathbf{x}_C | \mathbf{x}_R) = \frac{\tilde{p}_l(\mathbf{x}_C | \mathbf{x}_R)}{\int_{\tilde{\mathbf{x}}_C} \tilde{p}_l(\tilde{\mathbf{x}}_C | \mathbf{x}_R)} \quad (9)$$

It should be noted that \tilde{p}_l , as defined in Eq. (8), is a true probability only over d_l , i.e. over the orthogonal distance to the epipolar line. Since multiple points in 2D image space can have the same distance d_l from the epipolar line, Eq. (8) does not sum to one over this space. Therefore Eq. (8) is *not* a probability over the 2D image space and the normalization in Eq. (9) is required.

The completely analogous definition for p_m , the probability of a match with respect to motion, is omitted here for brevity. The standard deviations σ_l and σ_m should be set according to the expected error in calibration information.

Thus, for each possible correspondence point, two probabilities are computed: the probability that it is a good match according to its *location* (p_l ; this can be understood as a fuzzy version of the epipolar constraint), and the probability that it is a good match according to its *motion* (p_m). Further assuming that d_l and d_m are independent, a combined correspondence probability $p_{l,m}$ can be written as

$$p_{l,m}(\mathbf{x}_C | \mathbf{x}_R) = \frac{p_l(\mathbf{x}_C | \mathbf{x}_R) p_m(\mathbf{x}_C | \mathbf{x}_R)}{\int_{\tilde{\mathbf{x}}_C} p_l(\tilde{\mathbf{x}}_C | \mathbf{x}_R) p_m(\tilde{\mathbf{x}}_C | \mathbf{x}_R)} \quad (10)$$

In a similar fashion to Eq. (3), Eq. (10) can be refined taking a whole sequence (or period) of frames into account instead of only two subsequent points in time. Taking T frames into consideration and again using a weighting factor as in Eq. (3), the best correspondence $\tilde{\mathbf{x}}_C$ is computed as

$$\tilde{\mathbf{x}}_C = \operatorname{argmax}_{\mathbf{x}_C} \sum_{t=1}^T w(t, \mathbf{x}_R(t)) p_{l,m}(\mathbf{x}_C(t) | \mathbf{x}_R(t)). \quad (11)$$

Reconstruction results based on this will henceforth be referred to as *probabilistic reconstruction*.

3. RESULTS

The presented method is verified on one in vivo experiment obtained during an ongoing DIET clinical trial. Ethics approval for this study was obtained from the "Upper South A Ethics Committee", and written consent was obtained from all subjects. In order to allow a comparison of results from both the marker-based and the motion-based reconstruction, image sequences with attached markers were used. These images were first analyzed using the traditional, marker-based approach (the result of which are shown in the top row of Fig. 5). Afterwards, the markers were digitally removed from the images by skin-tone filtering, thus forcing the motion-based approaches to take only pure skin patches into account. Results are shown in Fig. 5.

The sampling distance used in the direct reconstruction was 4 samples per pixel, using bicubic interpolation of the optical flow information. The parameters used in the probabilistic reconstruction were $\sigma_l = 0.01$, $\sigma_m = 0.0025$.

The breast geometries reconstructed using marker-based, direct, and probabilistic reconstruction are shown in Fig. 5(a), Fig. 5(d), and Fig. 5(g), respectively. In this example, the number of tracked surface points was 1364 for the marker-based reconstruction, 3896 for the direct reconstruction and 3901 for the probabilistic reconstruction. The surface-to-surface error (i.e. the mean squared distance between points in the motion-based reconstructions and their respective projection onto the marker-based reconstruction) was 8.27 mm^2 for the direct reconstruction and 8.29 mm^2 for the probabilistic reconstruction.

For the actual diagnosis following the geometric reconstruction, the surface motion is characterized by a number of metrics. Since the breast moves in a steady state, each point on the surface moves on a closed, elliptical curve during one vibration period. The metrics describe properties such as the sizes, orientations, rotations and shapes of these ellipses. A number of these metrics is then used for the actual tumor classification.

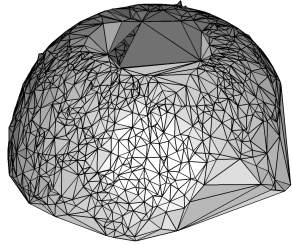
Fig. 5(b), Fig. 5(e) and Fig. 5(h) show one of the motion metrics, the *total displacement amplitude*, i.e. the L_2 -Norm of a vector containing the amplitudes of the periodic movements in all three dimensions. It is displayed on the reconstructed surface, viewed from above. In order to allow a classification of these surface motion metrics independent of the breast shape, the metrics are first projected onto a half-sphere. This half-sphere is then equally divided into segments in both ϕ (or longitudinal) and θ (or latitudinal) directions, and, for a given metric, the mean is computed for each segment. Fig. 5(c), Fig. 5(f) and Fig. 5(i) show the segment means for the total displacement amplitude for marker-based reconstruction, direct reconstruction and probabilistic reconstruction, respectively. Total errors for the segment means compared to the marker-based approach are shown in Fig. 6.

4. DISCUSSION AND CONCLUSIONS

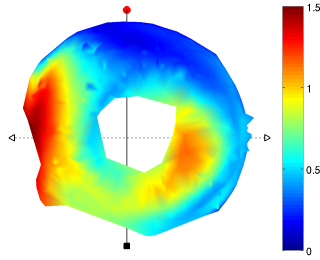
The DIET system relies on surface motion of an oscillating human breast. This is currently performed by tracking externally applied fiducial markers in 3D. To improve resolution and practical implementation shortcomings, a direct tracking of skin patterns is desired. A new, motion-based epipolar constraint approach was thus developed, that is able to achieve this task. Two alternative correspondence estimation techniques are presented, direct and probabilistic, and are validated in vivo.

Visual inspection of Fig. 5 shows that the results obtained using only motion are comparable to those relying on markers. This is not only the case for the geometric reconstructions, but also for the tracked motion of the surface, as can be seen in the amplitude of motion shown in Fig. 5(b), Fig. 5(e) and Fig. 5(h) and in the segment mean values. The problem of marker-less regions, as shown in the lower right of Fig. 5(a), where a large cutoff is visible, is solved.

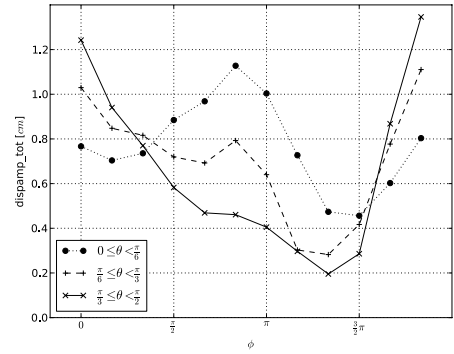
As given in the results, using motion-based reconstruction (either direct or probabilistic), the resolution of the reconstruction does not depend on the number of markers. Given sufficiently accurate motion information, an arbitrary large number of correspondences can be reconstructed, leading to meshes with very high resolutions. Furthermore, the finer resolution might provide means for additional accuracy in determining the presence and characteristics of a tumor within the breast.



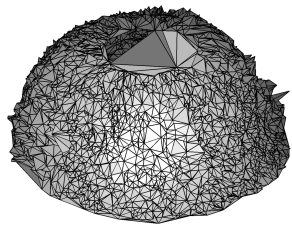
(a) Geometry reconstruction, marker-based approach



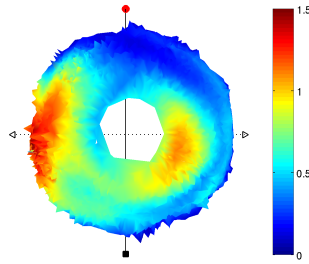
(b) Metric reconstruction, marker-based approach



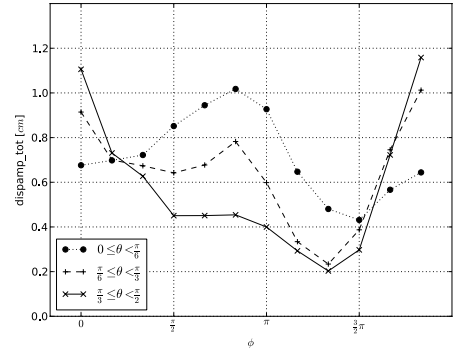
(c) Metric segment means, marker-based approach



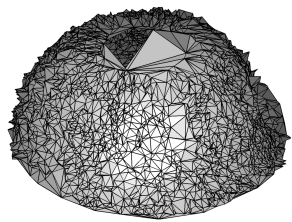
(d) Geometry reconstruction, motion-based approach (direct)



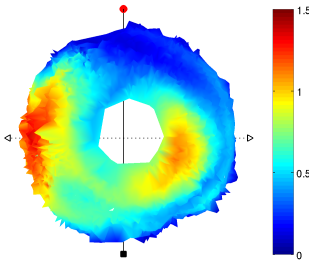
(e) Metric reconstruction, motion-based approach (direct)



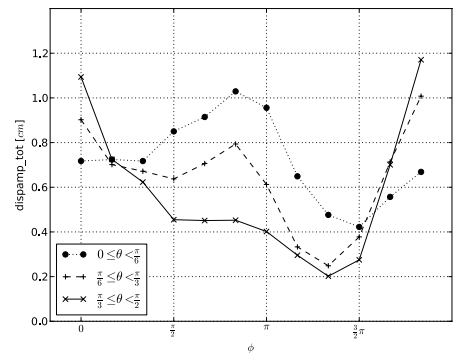
(f) Metric segment means, motion-based approach (direct)



(g) Geometry reconstruction, motion-based approach (probabilistic)

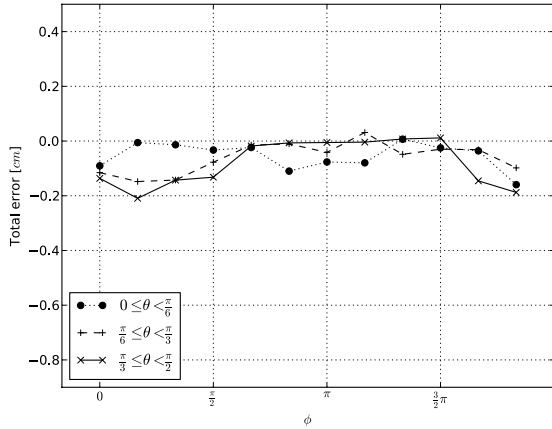


(h) Metric reconstruction, motion-based approach (probabilistic)

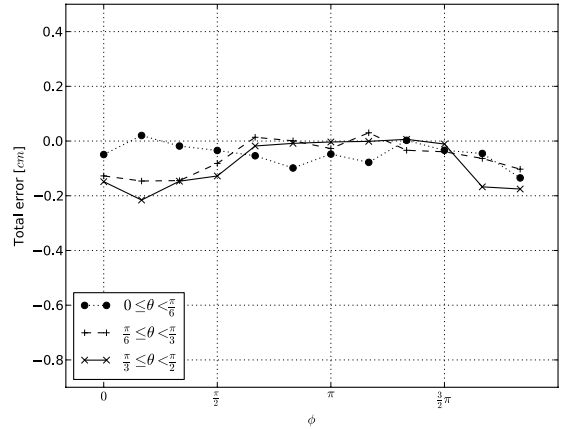


(i) Metric segment means, motion-based approach (probabilistic)

Figure 5: Reconstructions of breast geometry and surface motion. The surfaces are triangulated from the motion centerpoints.



(a) Segment mean errors, direct reconstruction



(b) Segment mean errors, probabilistic reconstruction

Figure 6: Segment mean errors

Fig. 5, Fig. 6, and the quantitative results given show that, in practical applications for the DIET system, direct reconstruction does not perform worse than the more complicated and computationally more expensive probabilistic reconstruction. However, as stated above, this is largely due to the availability of highly accurate calibration information (i.e. \mathbf{F}), and cannot be expected in general applications.

One drawback is the roughness of the surface, caused by small imprecisions in the optical flow computation. Further research on the possible benefits of different, more precise optical flow computation methods to overcome this are being evaluated. Additionally, the surfaces shown in Fig. 5 are created using simple triangulation from the reconstructed points and, as such, are linear interpolations. We expect a much better shape to result from non-interpolation methods, e.g. using the reconstructed points to fit a smooth model.

Furthermore, the methods presented here were only tested on a single dataset. More extensive validation on multiple datasets will be required to identify optimal values for the tunable parameters and robustness on varying skin types.

Overall, the presented method is able to robustly estimate correspondences from non-rigid motion information, in this case the surface of a vibrating breast. The improvements in handling and resolution are invaluable for the further development of the DIET system.

REFERENCES

- [1] Coughlin, S. S. and Ekwueme, D. U., “Breast cancer as a global health concern,” *Cancer epidemiology* **33**(5), 315–8 (2009).
- [2] Kriege, M., Brekelmans, C. T., Boetes, C., Besnard, P., and *et al.*, “Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition,” *The New England journal of medicine* **351**(5), 427–37 (2004).
- [3] Peters, A., Chase, J. G., and Van Houten, E. E., “Digital Image Elasto-Tomography: Combinatorial and Hybrid Optimization Algorithms for Shape-Based Elastic Property Reconstruction,” *IEEE transactions on bio-medical engineering* **55**(11), 2575–83 (2008).
- [4] Lotz, T., Simpson, P., Stocker, D., Hann, C., and Chase, J., “In Vitro Evaluation of Surface Based Non-Invasive Breast Cancer Screening with Digital Image based Elasto Tomography (DIET),” *Proceedings IEEE EMBS* (2010).
- [5] Brown, R., Hann, C., Chase, J., and Ray, L., “Discrete Colour-based Euclidean-Invariant Signatures for Feature Tracking in a DIET Breast Cancer Screening System,” *Proceedings SPIE Medical Imaging 2007* (2007).

- [6] Horn, B. K. and Schunck, B. G., “Determining Optical Flow,” tech. rep., Massachusetts Institute of Technology, Cambridge, MA, USA (1980).
- [7] Burger, W. and Bhanu, B., “Estimating 3D Egomotion from Perspective Image Sequence,” *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(11), 1040–1058 (1990).
- [8] Hatzitheodorou, M., Karabassi, E. A., Papaioannou, G., Boehm, A., and Theoharis, T., “Stereo Matching Using Optic Flow,” *Real-Time Imaging* **6**(4), 251–266 (2000).
- [9] Hartley, R. I. and Kang, S. B., “Parameter-Free Radial Distortion Correction with Centre of Distortion Estimation,” *IEEE Trans. Pattern Anal. Mach. Intell.* **29**, 1309–1321 (2007).
- [10] Dellaert, F., Seitz, S. M., Thorpe, C. E., and Thrun, S., “Structure from Motion without Correspondence,” in [*Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*], 557–564 (2000).
- [11] Lowe, D. G., “Distinctive Image Features from Scale-Invariant Keypoints,” *Int. J. Comput. Vision* **60**(2), 91–110 (2004).
- [12] Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L., “Speeded-Up Robust Features (SURF),” *Comput. Vis. Image Underst.* **110**(3), 346–359 (2008).
- [13] Morel, J.-M. and Yu, G., “ASIFT: A New Framework for Fully Affine Invariant Image Comparison,” *SIAM Journal on Imaging Sciences* **2**(2), 438–469 (2009).
- [14] Mitiche, A., “On Combining Stereopsis and Kineopsis for Space Perception,” in [*First International Conference on Artificial Intelligence Applications*], 156–160 (1984).
- [15] Mitiche, A., “Three-Dimensional Space from Optical Flow Correspondence,” *Computer Vision, Graphics, and Image Processing* **42**(3), 306–317 (1988).
- [16] Fermuller, C. and Aloimonos, Y., “On the Geometry of Visual Correspondence,” *Int. J. Comput. Vision* **21**(3), 223–247 (1997).